DOI:10.13203/j.whugis20190052 文章编号:1671-8860(2021)02-0303-06

# 抗翻录攻击的鲁棒语音水印算法

刘正辉[1,2] 张 钰[1] 秦兴红[2]

1 信阳师范学院计算机与信息技术学院,河南 信阳,464000
2 深圳大学信息工程学院深圳市媒体信息内容安全重点实验室,广东 深圳,518060

**摘 要**:水印为数字音频的版权保护提供了一种技术手段。然而,随着录用设备的普及,翻录攻击成为一种去除水印信息的有效方法。为了提高水印算法的安全性,提出了一种鲁棒的抗翻录攻击的数字语音水印算法。定义了离散余弦系数对数均值(discrete cosine transform coefficients logarithm mean,DCT-CLM)的特征,分析了该特征对翻录攻击的鲁棒性,并给出了基于该特征的水印嵌入方法。帧号和水印一起作为嵌入在各语音帧的信息,通过量化DCT-CLM方法将帧号和水印一起嵌入在各语音帧中。帧号用来同步各语音帧的内容,从同步的含水印语音帧中提取水印信息,从而进行溯源追踪。和常见的语音水印算法相比,该算法除了对去同步攻击的鲁棒性之外,还能够抵抗对敏感语音内容的翻录攻击。

**关键词**:数字语音;数字水印;去同步攻击;溯源追踪
**中图分类号**:P208 **文献标志码**:A

当前,人们可以使用多种设备来充分感受语音数据为大家提供的方便[1]。比较常见的有通过手机、平板电脑以及数码录音笔等录制自己感兴趣的音频信号,并通过网络传播、分享这些作品,以丰富人们的生活。然而,对于某些敏感的音频信号,这些传播势必会对作品所有者或社会带来消极的影响。对于非法传播者而言,要追究其责任。因此,对音频信号的溯源追踪成为当前研究的热点领域,尤其是对于敏感的语音信号。本文以数字语音为研究对象,以数字水印技术为方法,解决语音信号的溯源追踪问题。

近年来,已有学者对数字水印技术展开了研究[2-6]。由于音频信号的特殊性,去同步攻击会导致攻击后含水印音频的分帧和攻击前的不同步问题,使水印不能被正确地提取。为了抵抗去同步攻击,常用的方法是在音频各帧信号中嵌入标示信息。在水印检测时,首先检测各帧的标示信息,以定位相应帧的内容。文献[7]采用同步码的方法提出了一种鲁棒音频水印算法,通过量化离散余弦变换(discrete cosine transform,DCT)系数的方法来嵌入各帧的同步信息。相似地,文献[8]首先构造了16 bit的同步信息,然后将16 bit的同步信息采用量化伪Zernike矩的方法嵌入在每个音频帧的固定位置,以同步含水印的内容。

通常,频率域的音频水印算法具有较好的鲁棒性,常用于需要有较好的抗攻击能力的版权保护方面。当然,将水印嵌入在语音信号中,确实留下了可以用来追踪的信息[9-10]。但是,对于当前多数水印算法而言,翻录攻击后,很难在翻录的语音中提取到正确的水印信息[11-12]。这意味着当前多数的音频水印算法在语音信号的溯源追踪方面效果不显著。

考虑到当下对敏感数字语音溯源追踪的需求,本文提出了一种鲁棒的抗翻录攻击的语音水印算法。定义了离散余弦系数对数均值(discrete cosine transform coefficients logarithm mean,DCT-CLM)的特征,分析了该特征对翻录攻击的鲁棒性。将语音信号分帧,各帧的帧号映射为同步信息,和水印一起采用量化DCT-CLM特征的方法嵌入到各语音帧中。含水印语音在去同步攻击后,通过帧号可以使含水印内容再同步,进而提取有用的水印信息。实验分析了该算法的抗去同步攻击和翻录攻击的能力,验证了本算法对去同步攻击和翻录攻击的鲁棒性。

**收稿日期**:2019-09-12
**项目资助**:国家自然科学基金(61502409);深圳市媒体信息内容安全重点实验室开放基金(2018-05);信阳师范学院南湖学者奖励计划青年项目。
**第一作者**:刘正辉,博士,主要从事数字取证、信息隐藏研究。zhenghui.liu@163.com

# 1 DCT-CLM特征

## 1.1 DCT-CLM的定义

对语音信号$A$进行DCT后得到系数$C=\{c_n|1\leqslant n\leqslant N\}$，由系数$C$计算语音信号$A$的DCT-CLM特征，如下所示：

$$F=\left|\sum_{n=1}^{N}\log_2\frac{|c_n|}{\alpha}\right|/N \tag{1}$$

式中，$\alpha$为系统密钥，$\alpha>0$；$c_n$为第$n$个DCT系数，$c_n\neq 0$（若$c_n=0$，将$c_n$赋值为一个较小的值参与式(1)的计算，如将$c_n$赋值为0.000 1），且$|c_n|/\alpha<1$；$N$为DCT系数的个数，$1\leqslant n\leqslant N$。由式(1)可得，$C$中包含幅值大的系数越多，则DCT-CLM特征越小；相反，DCT-CLM特征越大。在鲁棒性分析中，也能看出DCT-CLM特征的该性质。

## 1.2 鲁棒性分析

一般而言，和音频信号相比，语音信号采样频率较低。随机选取一段长为$L$、采样频率为8 kHz的语音信号，记为$S$，如图1所示。步骤如下：

1)将$S$分为$P$帧，第$i$帧记为$S_i$，每帧长为$N$；

2)对$S_i$进行DCT，计算其DCT-CLM特征，记为$F_i$。

对语音信号进行翻录攻击，翻录后的信号记为$S'$，如图2所示。重复上述步骤1)和2)，将$S'$分帧，并计算各帧的DCT-CLM特征。图3给出了语音信号$S$和翻录攻击后信号$S'$各帧的DCT-CLM特征，其中每帧长$N$=500。

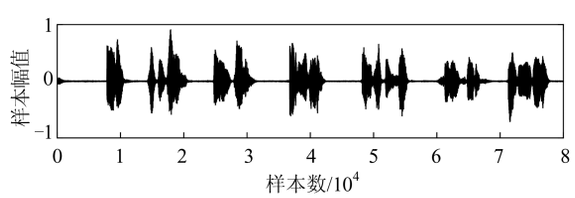


图1 原始语音信号

Fig.1 Original Speech Signal

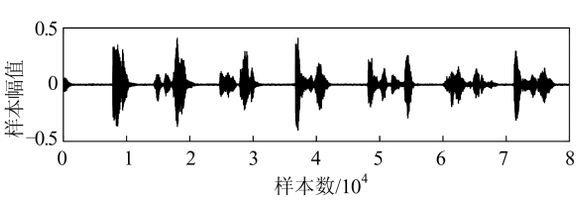


图2 翻录语音信号

Fig.2 Recaptured Speech Signal

为了测试DCT-CLM特征抵抗翻录攻击的普适性，选取300段不同的语音信号，并对选取的语音信号进行翻录攻击。按照上述方法，将各语音信号分为100帧。计算300段语音信号各帧攻击前后DCT-CLM特征的差值。图4给出了攻击前后各帧DCT-CLM特征差值的统计均值。可以看出，攻击前后DCT-CLM特征变化的最大幅度约为0.4，和DCT-CLM特征相比，变化幅度相对较小。若采用量化DCT-CLM特征的方法嵌入水印，可以保证含水印信号在翻录攻击后以较高的概率正确提取嵌入的信息。

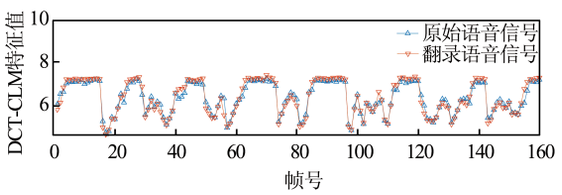


图3 原始语音信号和翻录语音信号的DCT-CLM特征

Fig.3 DCT-CLM Feature of Original and Recaptured Speech Signals

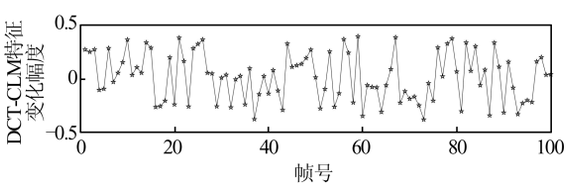


图4 原始语音信号翻录后DCT-CLM特征的变化幅度

Fig.4 Variation Amplitude of DCT-CLM Feature for Original Speech Signal After Being Recaptured

# 2 水印算法

记$W_i=\{w_1,w_2\cdots w_M\}$为第$i$帧要嵌入的信息，其中$w_m\in\{0,1\}$，$1\leqslant m\leqslant M$。将$W_i$分为3部分，分别记为$W_{1i}$、$W_{2i}$和$W_{3i}$，$W_{1i}=\{w_m|1\leqslant m\leqslant M_1\}$，$W_{2i}=\{w_m|M_1+1\leqslant m\leqslant 2M_1\}$，$W_{3i}=\{w_m|2M_1+1\leqslant m\leqslant M\}$。$W_{1i}$和$W_{2i}$分别由帧号生成，用来同步含水印的语音帧，这里$W_{1i}=W_{2i}$。$W_{3i}$为整个水印（版权）信息或者部分水印信息。在水印检测端，同样将一帧含水印的信号分为3部分。从前两部分中提取同步信息$W_{1i}$和$W_{2i}$。如果$W_{1i}=W_{2i}$，则表明该帧为同步的含水印内容，并从第3部分中提取水印信息。

## 2.1 水印嵌入

假设语音信号为$S=\{s_l|1\leqslant l\leqslant L\}$，其中$s_l$表示第$l$个样本点，$L$表示信号$S$的长度。步骤如下：

1)将语音信号$S$等分为$P$帧，第$i$帧记为$S_i$，每帧长为$N$，$N=L/P$。将$S_i$等分为$M$段，分别

为 $S_{i,1}$, $S_{i,2}\cdots S_{i,M}$,每段长为 $N_1$。

2)对 $S_i$ 的第 1 段 $S_{i,1}$ 进行 DCT,得到的系数为 $C_{i,1}=\{c_1, c_2\cdots c_{N_1}\}$。由式(1)计算 $S_i$ 各段的 DCT-CLM 特征,第 1 段的特征记为 $F_{i,1}$。

3)按照如下方法来量化特征 $F_{i,1}$:

$$QF_{i,1}=\begin{cases}\left\lfloor F_{i,1}/\Delta\right\rfloor\times\Delta+\Delta/2, \left\lfloor F_{i,1}/\Delta\right\rfloor \bmod 2=w_1\\ \left\lfloor F_{i,1}/\Delta\right\rfloor\times\Delta-\Delta/2, \left\lfloor F_{i,1}/\Delta\right\rfloor \bmod 2\neq w_1\end{cases} \tag{2}$$

式中,$QF_{i,1}$ 表示第 1 段量化后的 DCT-CLM 特征;$\lfloor\cdot\rfloor$表示向下取整;$\Delta$ 为量化步长。

4)由式(3)量化 $S_{i,1}$ 的 DCT 系数,$C_{i,1}=\{c_1, c_2\cdots c_{N_1}\}$,其中 $C^*_{i,1}=\{c^*_1, c^*_2\cdots c^*_{N_1}\}$ 表示量化后的系数。

$$c^*_n=\mathrm{sign}(c_n)\cdot|c_n|^{\frac{QF_{i,1}}{F_{i,1}}}\cdot\alpha^{1-\frac{QF_{i,1}}{F_{i,1}}}, 1\leqslant n\leqslant N_1 \tag{3}$$

式中,$\mathrm{sign}(c_n)$ 为符号函数。当 $c_n\geqslant 0$ 时,$\mathrm{sign}(c_n)=1$;当 $c_n<0$ 时,$\mathrm{sign}(c_n)=-1$。

5)对量化后的 DCT 系数 $C^*_{i,1}$ 进行逆 DCT,即可得到嵌入 $w_1$ 的语音信号。

重复上述步骤,可将 $w_2, w_3\cdots w_M$ 嵌入到 $S_i$ 其他各段 $S_{i,2}, S_{i,3}\cdots S_{i,M}$ 中。嵌入方法如图 5 所示。

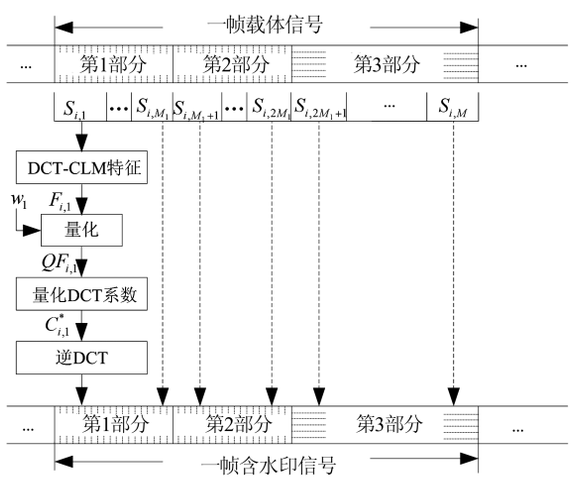


图 5　水印嵌入过程

Fig.5　Process of Watermark Embedding

## 2.2　水印提取

假设含水印信号为 $S^*$,长度为 $L^*$,水印提取步骤如下:

1)和水印嵌入过程相似,对含水印信号 $S^*$ 进行分帧,记第 $i$ 帧信号为 $S^*_i$。将 $S^*_i$ 等分为 $M$ 段,即 $S^*_{i,1}, S^*_{i,2}\cdots S^*_{i,M}$。

2)对 $S_{i,1}$ 进行 DCT,计算其 DCT-CLM 特征,记为 $F^*_{i,1}$。

3)由式(4)提取 $S^*_i$ 第 1 段 $S^*_{i,1}$ 中嵌入的信息 $w^*_1$:

$$w^*_1=\left\lfloor F^*_{i,1}/\Delta\right\rfloor \bmod 2 \tag{4}$$

重复步骤 2)和步骤 3),提取 $S^*_i$ 其他段中嵌入的信息 $w^*_2, w^*_3\cdots w^*_M$。记 $W^*_i=\{w^*_1, w^*_2\cdots w^*_M\}$,将 $W^*_i$ 分为 3 段,即 $W^*_i=[W^*_{1i}, W^*_{2i}, W^*_{3i}]$,这 3 段分别表示为:$W^*_{1i}=\{w^*_m|1\leqslant m\leqslant M_1\}$, $W^*_{2i}=\{w^*_m|M_1+1\leqslant m\leqslant 2M_1\}$, $W^*_{3i}=\{w^*_m|2M_1+1\leqslant m\leqslant M\}$。如果 $\sum_{m=1}^{M_1} w^*_m\oplus w^*_{m+M_1}=0$(⊕ 表示异或操作),表明该帧的内容是同步的,$W^*_{3i}$ 为提取的水印信息;如果 $\sum_{m=1}^{M_1} w^*_m\oplus w^*_{m+M_1}\neq 0$,则移动样本,将下一个连续的 $N$ 个样本分帧、分段,提取同步信息并重构帧号,直到从前两段提取的同步信息构造的帧号相等为止。至此,该连续的 $N$ 个样本即为找到的同步语音帧,可从该帧中提取水印信息。提取和同步水印的流程如图 6 所示。

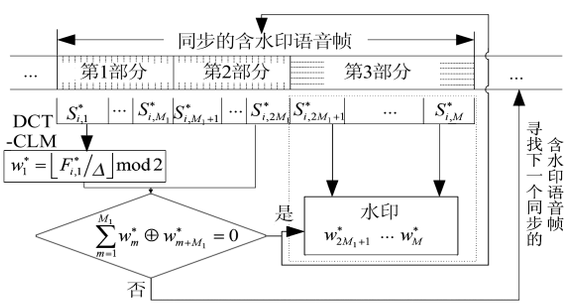


图 6　提取和同步水印的流程

Fig.6　Process of Watermark Extraction and Synchronization

对于去同步攻击而言,本文给出两类有代表性的攻击方法,分别记为 M1 和 M2。M1 是均匀地在含水印信号的每帧中插入等长的其他信号,M2 是在含水印信号的某一个位置插入其他信号,分别如图 7 和图 8 所示(为了便于说明,以插入 2 000 个样本点为例)。如果将含水印信号(被攻击信号)分为和水印嵌入端相同的分帧个数,则图 7 所示的信号被均匀地分为 10 帧,且每帧均包含 200 个其他的样本;而对于图 8 所示的情况而言,随着帧号的增加,每帧中包含攻击前对应帧的样本个数越来越少,这样攻击前后同一帧的内容相差较大,势必影响水印的提取。故对于图 8 所示的攻击类型,若采用和嵌入端帧长相等的分法,则攻击后的语音帧和攻击前完全相同(除了最后拼贴的信号之外),水印可以被正确地提取。

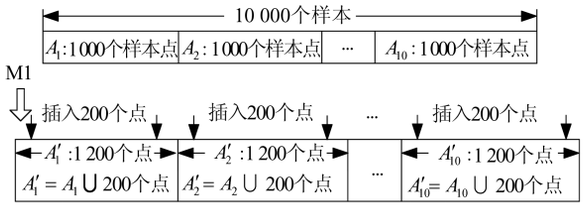


图7 第1种类型攻击M1分帧方法

Fig.7 Segmentation Method for the First Attack M1

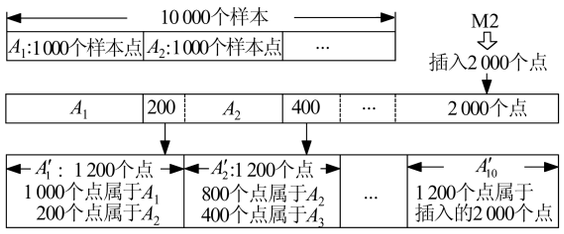


图8 第2种类型攻击M2分帧方法

Fig. 8 Segmentation Method for the Second Attack M2

基于以上分析,对待检测的含水印信号,尝试多种不同的长度(如90%×$L/P$、$L/P$、110%×$L/P$、$L^*/P$)进行分帧,以提取含水印语音中嵌入的信息,并提高水印提取的正确率。

## 3 性能测试

采用MATLAB对本文算法的综合性能进行仿真测试。选取300段不同类型的采样频率为8 kHz、16位量化的语音信号作为测试样本。对300段语音信号采用不同的录音设备进行翻录攻击,录音设备包括索尼录音笔(PCM-D100)、华为P20和iPhone 6s。实验用到的参数及取值分别为:帧长 $L=160\,000$, $N=8\,000$, $P=20$, $M=10$, $M_1=3$, $\Delta=0.8$, $\alpha=12$。

### 3.1 不可听性

水印的嵌入不应该影响原始信号的听觉质量,常用不可听性来测试水印的嵌入对原始信号听觉质量的影响程度。本文采用主观区分度(subjective difference grades, SDG)和信噪比(signal to noise ratio, SNR)[13]来测试算法的不可听性。

表1列出了对300段语音信号测得的SNR值和SDG值(最大值、最小值和均值),其中SDG值由11位听众根据原始信号和含水印信号听觉质量的差异,并根据SDG值评分标准现场打分所得。从表1的结果可知,本文算法的SNR值和SDG值均大于水印不可感知所要求的最小值,SNR值大于20,SDG值大于−1[13],表明本文水印的嵌入不影响原始信号的听觉质量。

**表1 含水印信号的SNR值和SDG值**

Tab.1 SNR and SDG Values of Watermarked Signal

| 指标 | 最大值 | 最小值 | 均值 |
|---|---|---|---|
| SNR | 33.28 | 27.52 | 30.73 |
| SDG | −0.84 | −0.41 | −0.65 |

### 3.2 鲁棒性

用于溯源追踪的水印算法应具有较好的抗攻击能力,在含水印信号被攻击后,应确保水印信息能被正确地提取。本文采用误码率(bit error rates, BER)来测试算法的鲁棒性,误码率越小,说明提取水印的错误率越低,算法的鲁棒性越好。BER的定义如下:

$$\mathrm{BER}=\frac{1}{M}\sum_{m=1}^{M} w(m)\oplus w^*(m) \tag{5}$$

式中,$M$表示嵌入水印的长度;$w(m)$和$w^*(m)$分别表示原始水印信息和提取的水印信息。

#### 3.2.1 信号处理和去同步攻击的鲁棒性

表2给出了对含水印信号进行信号处理和去同步攻击后的BER值。其中,信号处理包括添加高斯噪声(30 dB)、MP3压缩(64 kbit/s和128 kbit/s)、添加回声(40%)。去同步攻击包括抖动攻击,所选参数为1/10、1/100和1/1 000,表示每隔10个、100个和1 000个样本删除一个样本点;变速攻击,所选参数为80%、90%、110%、120%。

从表2的结果可得,本文算法在添加高斯噪声、128 kbit/s的MP3压缩以及每隔100个和1 000个样本删除一个样本点的抖动攻击中,水印提取BER值为0。而对于其他类型的信号处理和去同步攻击,虽然BER值有所增加,但低于文献[9]和文献[10]所给算法的测试结果。表明和文献[9]、文献[10]结果相比,本文算法具有更好的抗攻击能力。

#### 3.2.2 翻录攻击、信号处理和去同步攻击的鲁棒性

对含水印信号,首先采用不同的录音设备对其进行翻录攻击,然后对翻录后的信号进行信号处理和去同步攻击操作。表3给出了含水印信号在上述攻击后提取的BER值,其中信号处理和去同步攻击的类型和参数同表2。

从表3的结果可得,对于翻录攻击以及在翻录攻击的基础上进行的信号处理和去同步攻击,本文算法和文献[9]、文献[10]所给算法的水印提取BER值均有不同程度的增加。相比而言,对于单纯的翻录攻击,本文算法的BER值为4%,而文献[9]、文献[10]算法的BER值分别为15%和

18%。对于在翻录攻击基础上进行的其他攻击,本文算法的BER值同样低于文献[9]、文献[10]算法的BER值,说明本文算法对翻录攻击以及信号处理和去同步攻击具有更好的鲁棒性。

**表2　不同信号处理和去同步攻击后水印提取的BER值**

Tab.2　BER Values of Watermarking After Different Signal Processing Operations and De-synchronization Attacks

| 攻击类型 | | 参数 | BER/% | | |
|---|---|---|---|---|---|
| | | | 文献[9] | 文献[10] | 本文算法 |
| 信号处理 | 高斯噪声 | 30 dB | 7 | 8 | 0 |
| | 回声 | 40% | 5 | 3 | 1 |
| | MP3压缩 | 64 kbit/s | 5 | 4 | 1 |
| | | 128 kbit/s | 2 | 3 | 0 |
| 去同步攻击 | 抖动攻击 | 1/10 | 9 | 12 | 1 |
| | | 1/100 | 7 | 8 | 0 |
| | | 1/1 000 | 3 | 3 | 0 |
| | 变速攻击 | 80% | 16 | 18 | 8 |
| | | 90% | 12 | 16 | 5 |
| | | 110% | 9 | 11 | 3 |
| | | 120% | 14 | 15 | 6 |

注:嵌入率为10 bit/s,SNR取值约为30 dB

**表3　翻录攻击和不同信号处理以及去同步攻击后水印提取的BER值**

Tab.3　BER Values of Watermarking After Recapturing Attack, Different Signal Processing Operations and De-synchronization Attacks

| 攻击类型 | | 参数 | BER/% | | |
|---|---|---|---|---|---|
| | | | 文献[9] | 文献[10] | 本文算法 |
| 翻录攻击 | | | 15 | 18 | 4 |
| 信号处理 | 高斯噪声 | 30 dB | 17 | 22 | 4 |
| | 回声 | 40% | 16 | 20 | 5 |
| | MP3压缩 | 64 kbit/s | 17 | 19 | 6 |
| | | 128 kbit/s | 16 | 17 | 4 |
| 去同步攻击 | 抖动攻击 | 1/10 | 18 | 22 | 6 |
| | | 1/100 | 18 | 21 | 4 |
| | | 1/1 000 | 16 | 19 | 4 |
| | 变速攻击 | 80% | 25 | 29 | 11 |
| | | 90% | 22 | 26 | 9 |
| | | 110% | 21 | 24 | 6 |
| | | 120% | 23 | 26 | 8 |

注:嵌入率为10 bit/s,SNR取值约为30 dB

综合以上测试结果可知,本文算法的水印嵌入是不被听觉感知的,不影响原始语音信号的听觉质量;同时,含水印信号在经过不同类型的攻击后,水印能被准确地提取,表明本文算法具有较好的抗攻击能力。

## 4　结　语

基于敏感语音内容被翻录后肆意传播的问题,本文提出了一种用于溯源追踪的鲁棒语音水印算法。首先定义了语音信号的DCT-CLM特征;然后分析了该特征在翻录攻击后的变化幅度,得出DCT-CLM特征在翻录攻击前后变化量很小的结论。将帧号和水印一起作为要嵌入的信息,采用量化DCT-CLM特征的方法嵌入在每个语音帧中。含水印信号在去同步攻击后,帧号用来同步被攻击的语音帧。在检测到同步的语音帧后,从同步的语音帧中提取水印信息,用来溯源追踪。和现有的语音水印算法相比,本文算法不仅提高了去同步攻击的水印提取正确率,而且具有抵抗翻录攻击的能力。

## 参　考　文　献


[1] Peng Zhenghong, Sun Zhihao, Cheng Qing, et al. Urban Land Use Function Recognition Method Using Sequential Mobile Phone Data[J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(9):1 399-1 407(彭正洪,孙志豪,程青,等. 利用时序手机通话数据识别城市用地功能[J]. 武汉大学学报·信息科学版,2018,43(9):1 399-1 407)

[2] Hua G, Huang J W, Shi Y Q, et al. Twenty Years of Digital Audio Watermarking—A Comprehensive Review[J]. *Signal Processing*,2016,128(11):222-242

[3] Hou Xiang, Min Lianquan. A Robust Watermarking Algorithm Using SURF Feature Regions[J]. *Geomatics and Information Science of Wuhan University*,2017,42(3):421-426(侯翔,闵连权. 基于SURF特征区域的鲁棒水印算法[J]. 武汉大学学报·信息科学版,2017,42(3):421-426)

[4] Nishimura A. Audio Watermarking Based on Subband Amplitude Modulation[J]. *Acoustical Science and Technology*,2010,31(5):328-336

[5] Lin Wei, Wang Yuhai, Ren Na, et al. QR Code Based Research on Digital Watermarking Algorithm for Tile Remote Sensing Image[J]. *Geomatics and Information Science of Wuhan University*, 2017, 42(8):1 151-1 158(林威,王玉海,任娜,等. 基于QR码的瓦片遥感影像数字水印算法[J]. 武汉大学学报·信息科学版,2017,42(8):1 151-1 158)

[6] Hou Xiang, Min Lianquan, Tang Liwen. Fragile Watermarking Algorithm for Locating Tampered Entity Groups in Vector Map Data[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45

(2):309-316(侯翔,闵连权,唐立文.定位篡改实体组的矢量地图脆弱水印算法[J].武汉大学学报·信息科学版,2020,45(2):309-316)

[7] Xiang Y,Natgunanathan I,Guo S,et al. Patchwork-Based Audio Watermarking Method Robust to De-synchronization Attacks[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2014,22(9):1 413-1 423

[8] Wang X Y,Ma T X,Niu P P. A Pseudo-Zernike Moments Based Audio Watermarking Scheme Robust Against Desynchronization Attacks[J]. *Computers and Electrical Engineering*, 2011, 37(4): 425-443

[9] Kang X G,Yang R,Huang J W. Geometric Invariant Audio Watermarking Based on an LCM Feature[J]. *IEEE Transactions on Multimedia*, 2011, 13(2): 181-190

[10] Nadeau,Sharma G. An Audio Watermark Designed for Efficient and Robust Resynchronization After Analog Playback[J]. *IEEE Transactions on Information Forensics and Security*,2017,12(6) 1 393-1 405

[11] Natgunanathan I, Xiang Y, Hua G. Patchwork-Based Multilayer Audio Watermarking[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*,2017,25(11):2 176-2 187

[12] Hu H T,Hsu L Y. Robust, Transparent and High-Capacity Audio Watermarking in DCT Domain[J]. *Signal Processing*,2015,109(3):226-235

[13] Liu Z H, Zhang F, Wang J, et al. Authentication and Recovery Algorithm for Speech Signal Based on Digital Watermarking[J]. *Signal Processing*,2016, 123(1):157-166

# Robust Speech Watermarking Algorithm Against Recapturing Attacks

*LIU Zhenghui*[1,2] *ZHANG Yu*[1] *QIN Xinghong*[2]

1 College of Computer and Information Technology, Xinyang Normal University, Xinyang 464000, China

2 Shenzhen Key Laboratory of Media Security, College of Information Engineering, Shenzhen University, Shenzhen 518060, China

**Abstract**: Watermarking provides a technical mean for copyright protection of digital audio. However, with the popularity of recording equipment, recapturing attack has become an effective method to remove audio watermarks. In order to improve the security of the watermarking system, we propose a robust speech watermarking algorithm against recapturing attacks. Firstly, we define the discrete cosine transform coefficients logarithm mean (DCT-CLM) feature and get the conclusion that the changes of DCT-CLM feature are very small after recapturing attacks. Secondly, Frame number and watermark are embedded together in frames by quantifying the DCT-CLM feature. Frame number is used to resynchronize watermarked speech after the signal is subjected to de-synchronization attacks. If watermarked frame is synchronized, we extract watermark bits from the frame for resource tracing. Compared with other speech watermarking algorithms, the algorithm proposed in this paper is not only robust against de-synchronization attacks, but also robust against recapturing attacks.

**Key words**: digital speech; digital watermarking; de-synchronization attacks; resource tracing

**First author:** LIU Zhenghui, PhD, specializes in digital forensics and information hiding. E-mail:zhenghui.liu@163.com

**Foundation support:** The National Natural Science Foundation of China (61502409);the Open Fund of Shenzhen Key Laboratory of Media Security (2018-05);Nanhu Scholars Program for Young Scholars of Xinyang Normal University.

**引文格式**:LIU Zhenghui,ZHANG Yu,QIN Xinghong.Robust Speech Watermarking Algorithm Against Recapturing Attacks[J].Geomatics and Information Science of Wuhan University,2021,46(2):303-308.DOI:10.13203/j.whugis20190052(刘正辉,张钰,秦兴红.抗翻录攻击的鲁棒语音水印算法[J].武汉大学学报·信息科学版,2021,46(2):303-308.DOI:10.13203/j.whugis20190052)